R E P O R T R E S U M E S

ED 003 051 24

MULTIVARIATE STATISTICAL PROCEDURES IN PREDICTING TEACHER-PUPIL CLASSROOM BEHAVIOR.

BY- SOAR, ROBERT S.
SOUTH CAROLINA UNIV., COLUMBIA

REPORT NUMBER CRP-1170 PUB DATE 62
EDRS PRICE MF-$0.27 HC-$7.24 181P.

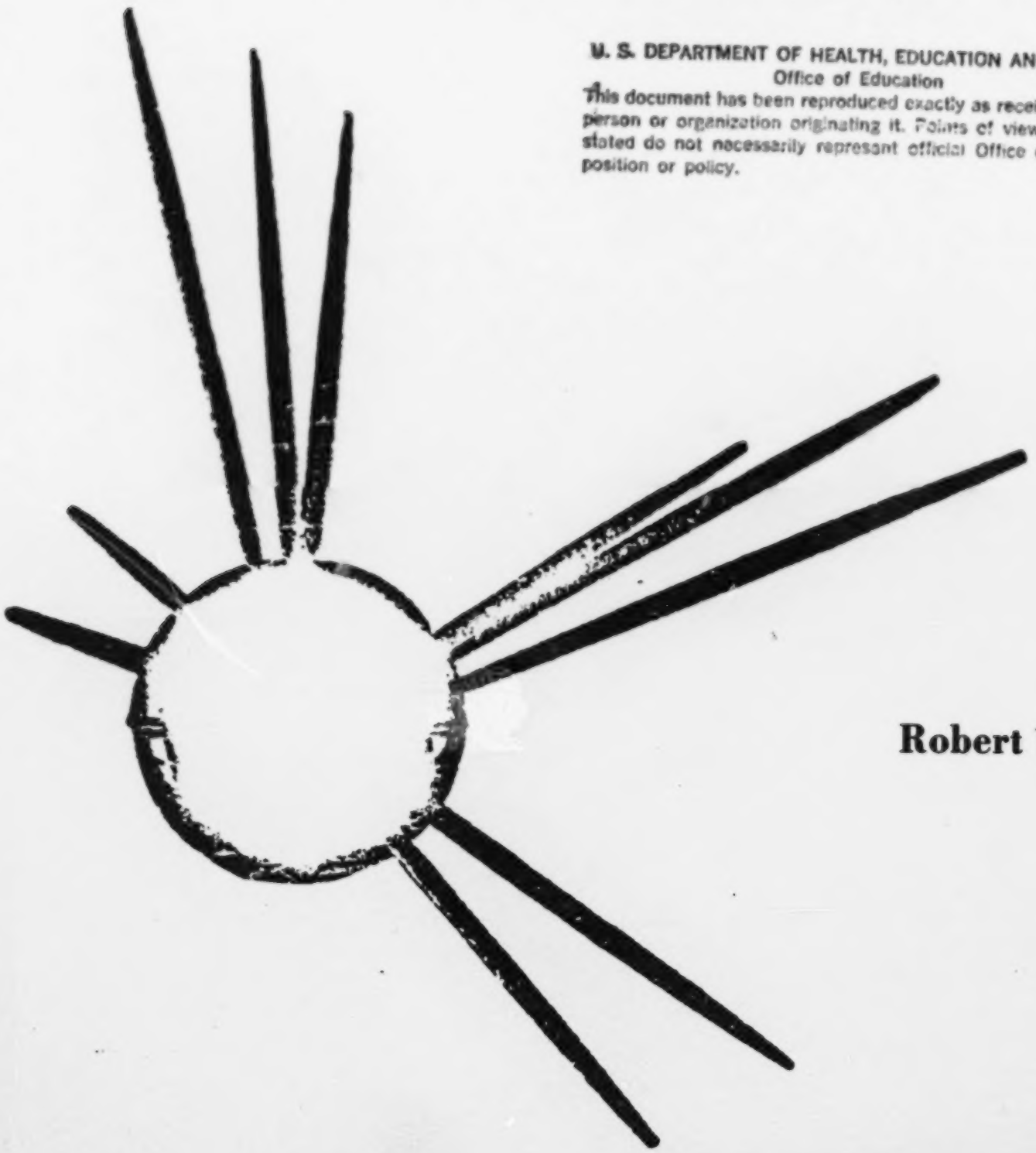
DESCRIPTORS- *FACTOR ANALYSIS, EVALUATION METHODS, *TEACHER RATING, TEACHER CHARACTERISTICS, TEACHER ATTITUDES, MEASUREMENT TECHNIQUES, *BEHAVIOR RATING SCALES, STUDENT BEHAVIOR, STUDENT TEACHER RELATIONSHIP, *CLASSROOM ENVIRONMENT, *PREDICTIVE ABILITY (TESTING), ELEMENTARY SCHOOL TEACHERS, EFFECTIVE TEACHING, COLUMBIA, SOUTH CAROLINA

NEWLY DEVELOPED MEASURES OF CLASSROOM BEHAVIOR AND THE DETERMINATION OF THEIR EFFECTIVENESS IN PREDICTING EFFECTIVE TEACHER-PUPIL INTERACTION WERE INVESTIGATED. THE MEASURES OF TEACHER CHARACTERISTICS WERE CORRELATED WITH PUPIL CHARACTERISTICS FROM WHICH ANALYSES IT WAS CONCLUDED THAT NO DEFINITE CONCLUSIONS COULD BE DRAWN NOR COULD REASONS BE FOUND TO EXPLAIN THE DIFFERENCES FROM SAMPLE TO SAMPLE. THE AUTHOR STATES THAT THERE IS NEED FOR FURTHER RESEARCH INTO THE METHODOLOGY OF TEACHER EFFECTIVENESS RESEARCH. (JK)

MULTIVARIATE STATISTICAL PROCEDURES IN PREDICTING TEACHER - PUPIL CLASSROOM BEHAVIOR

Cooperative Research Project No. 1170

U. S. DEPARTMENT OF HEALTH, EDUCATION AND WELFARE
Office of Education
This document has been reproduced exactly as received from the person or organization originating it. Points of view or opinions stated do not necessarily represent official Office of Education position or policy.



Robert S. Soar

University of South Carolina
Columbia, South Carolina

ED003051

MULTIVARIATE STATISTICAL PROCEDURES IN PREDICTING TEACHER-PUPIL CLASSROOM BEHAVIOR

Cooperative Research Project No. 1170

Robert S. Soar

University of South Carolina
Columbia, South Carolina

1962

The research reported herein was supported through the Cooperative Research Program of the Office of Education, U. S. Department of Health, Education, and Welfare.

3414w
116

ACKNOWLEDGMENTS

Without the willingness of classroom teachers to take on still another task in the hope that their profession will be advanced and pupils benefitted, this project would not have been possible. Their dedication has been an ever-present inspiration.

The contribution of the Nashville and Davidson County, Tennessee, school systems, both teachers and administrative staff, was basic to the earlier study from which a portion of the data was drawn. They were acknowledged in the report of that study, but should be recognized again here. In the South Carolina data collection particular thanks are due to Claud E. Kitchens and W. J. Castine for assistance in getting underway, and to Curry McArthur and Robert Newton of Sumter for backstopping at a last critical moment.

Norman D. Bowers was a colleague in the preceding study from which this one stemmed, and participated in the initial planning of this one. Mrs. Nancy Norris and Gabriel Batarseh carried out the classroom observations, and Mrs. Nancy Beskid and Mrs. Louise Dawkins administered the situational tests of pupil social skill. Both teams regularly worked beyond any reasonable expectation. Miss Bobbie Gibbs and Miss Charlotte Gibbs carried out data reduction and analysis with unusual responsibility, and Miss Patricia Leggett has brought continuing care and attention to detail in scoring, analyses and stenographic tasks. Mrs. Edna Spears has helped in time of need from beginning to end.

Sam C. Webb of Emory University consulted on the application of factor analytic procedures and supervised the processing of the data. Howard Rolf of the Vanderbilt University Computer Center "debugged" the canonical

correlation program and carried out the computations. J. Hubert Noland, Director of the University of South Carolina Computer Center, has given numerous hours of assistance in data processing. Dr. William W. Savage, Dean, School of Education, has provided freedom for inquiry and administrative support which has saved much lost effort and smoothed the way.

Very special appreciation is due my wife, Ruth M. Soar, who has not only provided support and encouragement but has participated as a colleague in every sense of the word. She has puzzled out new analyses from abstruse descriptions, organized both data and procedures for all of us, contributed more than a fair share of insights into sources of difficulty and the meaning of data, and shared the labor of writing. Without her skills this product must have been much delayed and much less adequate.

To all, my heartfelt appreciation.

Robert S. Soar
Columbia, S. C.

June, 1962

TABLE OF CONTENTS

		Page
LIST OF TABLES		vii
Chapter		
I.	INTRODUCTION	1
	Background of the Study	1
	Significance of the Study	2
II.	REVIEW OF THE LITERATURE	6
	Methodological Problems	6
	Criterion Problems	6
	Definition of a Criterion	6
	Measurement Problems	8
	Criterion Complexity	10
	Pupil Change Criteria	10
	Classroom Process Criteria	11
	A Proposed Framework	12
	Development of Composites	13
	Problems of General Methodology	14
	Choice of Subjects	14
	Failure to Consider Pupil Differences	14
	Summary	15
	Research Results	16
	Correlates of Teacher Effectiveness	16
	Factor Analyses of Teacher Behavior	17
	Summary	24
III.	PROCEDURE	25
	Objectives	25
	Subjects	26
	The First Sample	26
	The Second Sample	27
	Comparison of Samples	28
	Measures	29
	Bowers Teacher Opinion Inventory (BTOI)	29
	Minnesota Teacher Attitude Inventory (MTAI)	30
	The Survey of Educational Leadership Practices (Survey)	31
	The Minnesota Multiphasic Personality Inventory (MMPI)	32
	Special Keys	32
	Anxiety (A)	32
	Repression (R)	32

TABLE OF CONTENTS (Continued)

Chapter		Page
	Ego Strength (ES)	33
	Hostility, Pharisaic Virtue and Teacher Attitude (Ho, PV, and TA)	33
	Research Relating the MMPI to Teacher Effectiveness	34
	Factor Analytic Studies of the MMPI	35
	The Observation Schedule and Record (OSCaR)	37
	Russell Sage Social Relations Test (RSSR)	39
	Data Collection and Reduction	42
	Analysis	43
	Criterion Composites	44
	By Canonical Correlation	44
	By Scaling by Reciprocal Averages	45
	By Averaging Standard Scores	45
	Comparison of Composites	45
	Comparison of Canonical and Multiple Correlations	46
	Item Analysis	47
	Factor Analysis	48
	Cross-validation	48
	Summary	50
IV.	RESULTS AND DISCUSSION	51
	Comparison of Samples	51
	First Sample	54
	Validity Coefficients	54
	Classroom Behavior Composites	55
	Canonical Correlation	57
	Scaling by Reciprocal Averages	59
	Average of Standard Scores	64
	Comparison of Composites with Criterion Measures	64
	Comparison of Composites With Each Other	65
	Comparison of Composites with Teacher Characteristics.	67
	Comparison of Canonical and Multiple Correlations	69
	Factor Analysis	73
	Item Analysis	79
	Second Sample	80
	Comparison of the Samples	80
	Hypotheses for the Differences	81
	Criterion Inconsistency.	81
	Chance Relationships	84
	Sample Differences	86
	Motivational Differences	89
	Cross-validation of Weighting Schemes	93
	Cross-validated Canonical Correlations	93
	Cross-validated Multiple Correlations	93
	Cross-validated Item Analyses	95

TABLE OF CONTENTS (Continued)

Chapter		Page
	Comparison of Composites	98
	Factor Analysis	103
	Summary and Implications	106
	Non-replication and its Implications	106
	Comparisons of Composites	109
	Results of Predictive Procedures	110
	Results of Factor Analyses	110
V.	SUMMARY AND CONCLUSIONS	114
	Background	114
	Objectives	114
	Procedure	115
	Sample	115
	Measures	116
	Analyses	117
	Results	118
	Conclusions and Implications	120
	Classroom Behavior Composites	120
	Cross-validation	120
	Factor Analysis	123
	Implications	124
REFERENCES		125
APPENDIX A	Personal and Professional Characteristics of the Teachers	130
APPENDIX B	Scoring Procedure for the Observation Schedule and Record	135
APPENDIX C	Empirical Weights for the Russell Sage Social Relations Test	141
APPENDIX D	Hand Method for Scaling by Reciprocal Averages	143
APPENDIX E	Probit Analysis of Reciprocal Averages Equal Frequency Scores	146
APPENDIX F	Intercorrelations Between Attitude and Personality Measures and Classroom Behavior Measures for Two Samples of Elementary Teachers	148
APPENDIX G	Additional Factor Analyses (31 Variables, First Sample)	155
APPENDIX H	Extreme Groups Selected by Different Composites and MMPI Scoring Keys for Two Samples of Elementary Teachers	160

LIST OF TABLES

Table		Page
1.	Correlations of Teacher Attitude and Personality Measures with Measures of Classroom Behavior for Fifty-four Elementary Teachers (First Sample)	52
2.	Correlations of Teacher Attitude and Personality Measures with Measures of Classroom Behavior for Eighty-eight Elementary Teachers (Second Sample).	53
3.	Intercorrelations of Classroom Behavior Measures for Fifty-four Elementary Teachers (First Sample).	56
4.	Canonical Correlation and Combining Weights for Classroom Behavior Measures and Ten Attitude and Personality Measures for Fifty-four Elementary Teachers (First Sample)	58
5.	Canonical Correlation and Combining Weights for Classroom Behavior Measures and Ten Personality Measures for Fifty-four Elementary Teachers (First Sample)	60
6.	Canonical Correlation and Combining Weights for Classroom Behavior Measures and Four Personality Measures for Fifty-four Elementary Teachers (First Sample)	60
7.	Combining Weights By Scaling By Reciprocal Averages Based on Equal Intervals for Fifty-four Elementary Teachers (First Sample)	62
8.	Combining Weights by Scaling by Reciprocal Averages Based on Equal Frequencies for Fifty-four Elementary Teachers (First Sample)	63
9.	Intercorrelations Between Composites of Classroom Behavior Measures and Criterion Measures for Fifty-four Elementary Teachers (First Sample)	66
10.	Intercorrelations of Composites of Classroom Behavior Measures for Fifty-four Elementary Teachers (First Sample)	66
11.	Intercorrelations Between Composites of Classroom Behavior Measures and Four Personality Measures for Fifty-four Elementary Teachers (First Sample)	68
12.	Multiple Correlations and Combining Weights for Classroom Behavior Measures and Four Personality Measures for Fifty-four Elementary Teachers (First Sample)	70

LIST OF TABLES (Continued)

Table		Page
13.	Multiple Correlations and Combining Weights for Classroom Behavior Measures and Selected Attitude and Personality Measures for Fifty-four Elementary Teachers (First Sample)	72
14.	Factor Loadings by Varimax Rotation for Predictor and Criterion Measures for Fifty-four Elementary Teachers (First Sample)	74
15.	Minnesota Multiphasic Personality Inventory Items Which Discriminated Upper and Lower Groups Using Different Criterion Composites and Cutting Points	80
16.	Intercorrelations of Factor Scores and Sub-keys of the Observation Schedule and Record for Fifty-four Elementary Teachers (First Sample)	82
17.	Intercorrelations of Factor Scores and Sub-keys of the Observation Schedule and Record for Eighty-eight Elementary Teachers (Second Sample)	83
18.	Intercorrelations of Classroom Behavior Measures for Eighty-eight Elementary Teachers (Second Sample)	85
19.	Comparison of Means and Standard Deviations of Attitude and Personality Measures for Two Groups of Elementary Teachers	88
20.	Correlations Between Attitude and Classroom Behavior Measures for Experimental and Control Groups Separately (First Sample)	90
21.	Correlations Between Personality and Classroom Behavior Measures for Experimental and Control Groups Separately (First Sample).	91
22.	Comparison of Validity Coefficients for Experimental and Control Groups (First Sample)	92
23.	Correlations of Teacher Attitude and Personality Measures with Measures of Classroom Behavior for a Subgroup of Elementary Teachers (Second Sample, N = 39)	94
24.	Relationships Between Composites of Minnesota Multiphasic Personality Inventory and Classroom Behavior Measures Cross-validated on Eighty-eight Elementary Teachers . . .	95
25.	Relationships Between Composites of Attitude and Personality Measures and Classroom Behavior Measures Cross-validated on Eight-eight Elementary Teachers	96

LIST OF TABLES (Continued)

Table		Page
26.	Cross-validation of Four Keys Developed for the Minnesota Multiphasic Personality Inventory by Item Analysis	97
27.	Intercorrelations Between Composites of Classroom Behavior Measures and Four Criterion Measures for Eight-eight Elementary Teachers (Second Sample)	99
28.	Combining Weights by Scaling by Reciprocal Averages Based on Equal Frequencies for Ninety Elementary Teachers (Second Sample)	100
29.	Intercorrelations of Composites of Classroom Behavior Measures for Eighty-eight Elementary Teachers (Second Sample) . . .	101
30.	Intercorrelations Between Composites of Classroom Behavior Measures and Four Personality Measures for Eighty-eight Elementary Teachers (Second Sample)	102
31.	Factor Loadings by Varimax Rotation for Predictor and Criterion Measures for Eighty-eight Elementary Teachers (Second Sample)	103a
32.	Grade Taught by the Teachers	131
33.	College Degrees Held by the Teachers	131
34.	Sex of the Teachers	132
35.	Years of Teaching Experience	132
36.	Age of the Teachers	133
37.	Sex of the Teachers	133
38.	Education of the Teachers	134
39.	Probit Analysis of Reciprocal Averages Equal Frequency Scores (First Sample)	147
40.	Intercorrelations Between Attitude and Personality Measures and Classroom Behavior Measures for Fifty-four Elementary Teachers (First Sample)	149
41.	Intercorrelations Between Attitude and Personality Measures and Classroom Behavior Measures for Eighty-eight Elementary Teachers (Second Sample)	153

LIST OF TABLES (Continued)

Table		Page
42.	Factor Loadings by Varimax Rotation for Predictor and Criterion Measures for Fifty-four Elementary Teachers (First Sample) .	156
43.	Factor Loadings by Patterned Quartimax Rotation for Predictor and Criterion Measures for Fifty-four Elementary Teachers (First Sample)	158
44.	Comparison of Extreme Groups Selected by Four Composites (First Sample)	161
45.	Comparison of Extreme Groups Selected by Three Composites (Second Sample).	163
46.	Items Selected by Item Analysis of MMPI Using Two Criterion Composites (First Sample)	166
47.	Items Selected by Item Analysis of MMPI Using Reciprocal Averages Composite (Second Sample)	167

CHAPTER I

INTRODUCTION

Background of the Study

This project is an outgrowth of an earlier one carried out by the investigator and a colleague to study a somewhat different problem (Bowers and Soar, 1961). The earlier project was concerned primarily with evaluating the effectiveness of a workshop experience for elementary teachers intended to increase skill in working with pupil groups in particular, and inter-personal skill in general. As a part of that study, criterion data were collected which specified relatively objectively a number of dimensions of the classroom in which change might occur. A variety of personality and attitudinal teacher characteristics which might influence both initial teacher classroom behavior, and teacher change as a consequence of the laboratory experience were also measured along with a measure of the group problem solving skills of each pupil group. It is important in this context to note that the criterion measures, in addition to being relatively objective, were multidimensional, sampling several aspects of the differences from classroom to classroom.

In the course of the analysis of data toward testing the hypotheses for which the study was designed, it became clear that there were numerous significant relationships between the antecedent teacher characteristics and the classroom measures. And although these were not the goals of the original research, they appeared to be worthy of further study.

Significance of the Study

One of the aspects of the criterion data which argued for its further study was the fact that it skirted the value question of "good-poor" categorization of teacher behavior. This offered several advantages. First, rater differences in judgments about the desired outcomes of education were not involved. As will be discussed in Chapter II, this problem is one on which numbers of previous researches have foundered. Second, this kind of criterion specification avoids the implicit assumption that good or poor teaching is a unitary performance, and explicitly substitutes a multidimensional approach which the method of development suggests should summarize at least several aspects of classroom variance.

This approach to the study of the relationship between teacher characteristics and classroom behavior offers both pragmatic and theoretical advantages. Pragmatically, definitions of "good" teaching will differ from school to school and system to system. There will be some schools in which the paramount constellation of values are those that Thelen refers to as "toilet training"— quiet, order, conformity, and respect for authority--whereas in other schools the goals of education will be more broadly conceived. Significant relationships between measures of teacher characteristics and measures of classroom behavior would then permit identification of teachers or teacher trainees who would be likely to fit the expectations of a given school, system, or teacher education program.

Theoretical advantages follow from conceptualizing the research problem as one with at least three phases: (1) measurement of teacher characteristics, (2) measurement of classroom behavior of teacher and pupils, and, (3) measurement of pupil change and growth. If the research is based

on theory and the process specified in these steps, the functional relationships linking each phase to the next can be examined by research, and should be meaningful. When the jump is taken from teacher characteristics to pupil learning, as has sometimes been done in previous research, the interpretation of the findings in terms of existing theory is often difficult if not impossible; and the results cannot be easily used to generate new theory or extend current theory because of the uncertainty of the intervening processes. Each of the links, however, is a point in the over-all process at which behavior theory should be relevant in explaining the relationships observed. That is, it should be possible to relate teacher characteristics to teacher behavior by theory; similarly, it should be possible to relate pupil learning to teacher-pupil classroom behavior by theory; but to make the jump from teacher characteristics to pupil learning involves so many unknowns in the intervening process that relevance to theory becomes quite uncertain, even though such results might be used for empirical predictive purposes.

The study to be presented here is concerned primarily with the link between teacher characteristics and teacher-pupil behavior. This approach is only likely to contribute to understanding of the educational process if the dimensions of the classroom measured are sufficiently comprehensive to give promise of describing important differences between classrooms, and if these differences are ones which fit into a theory of pupil learning. With an appropriate base in theory, research may fruitfully proceed alternatively on either link in the over-all process, with real hope of integration of the results.

With the recognition of the multidimensionality of the phenomena being studied, further problems arise. One of these is the problem of how to

combine the various dimensions into a meaningful (or reliable, or predictable) whole. If the dimensions of teacher behavior in the classroom (or more accurately, the teacher-pupil behavior that occurs there partially in response to the behavior of the teacher), are not highly correlated, then it will be unlikely that all of the desirable behaviors will be found in one person very frequently, so that the problem of what weight to give each dimension in arriving at an over-all evaluation will be raised. Perhaps ideally this problem should be settled as a value judgment, but past attempts to reach agreement have not been successful and research on teacher effectiveness has not been advanced. An alternative approach would be to examine several statistical procedures, in order to gain some idea of the nature and relative utility of the composites which result.

Historically, science has often progressed by first reducing its data to the basic elements (analysis), and then combining or recombining these elements into a whole (synthesis). Recently educational research has begun to make progress with the analysis problem, but very little progress has been made with the problem of synthesis.

There are several possible statistical procedures which seem worth exploring. The simplest would be to average standard scores so that equal weights are assigned. Another would be to create the most reliable composite of teacher-pupil behavior; another would be to create the composite that maximizes the relation between teacher characteristics and classroom behavior; and still another would be to seek the underlying dimensions basic to both sets of data, and examine the way in which each set relates to these dimensions. Although attempts have been made to identify underlying dimensions of personality by factor analyses of objective measures, there have been few attempts to find common dimensions underlying teacher

personality and classroom behavior, using objective measures of both. If any of these several procedures produced similar results, they could be applied with greater confidence.

The feasibility of such a project has increased sharply over the past few years as a consequence of the increasing availability of electronic computers. The number of relevant variables which should be screened is large, and the screening process is economically carried out only by this means. In addition, the classical, single variable model taken over from the physical sciences is increasingly being called into question as adequate for behavioral research. It assumes "all else equal", and in settings where this is never true, and where interactions between variables are often the critical determinants of results, the importance of examining simultaneously a large number of aspects of behavior is being more and more widely accepted. This, too, is really economic only through the use of computers.

It seems reasonable to expect that material advances in educational research in particular, and in behavioral research in general, will be made as a consequence of the increased ability to deal simultaneously with numbers of elements of a behavioral complex such as a classroom.

Stemming from an earlier study, then, this is a further study of the relations between a number of antecedent teacher characteristics and the teacher-pupil classroom behaviors associated with them. The nature of the criterion composites that result when different statistical procedures are applied will be examined, along with the predictability of these composites and the characteristics of teachers which are associated with them. The underlying dimensions will be sought by factor analysis, and an attempt will be made to develop a scale from the teacher personality measures which will predict the composites. Finally, a new sample of data will be collected, and the various procedures cross-validated.

CHAPTER II

REVIEW OF THE LITERATURE

Despite an extensive literature on teacher effectiveness which has accumulated over the years, one of the central points on which reviewers agree is that very limited progress has been made toward prediction and assessment (Tiedeman and Cogan, 1958; Levin, 1954; Mitzel, 1960; AASA, 1961; Remmers et al, 1952, 1953). The number of studies in which objective procedures for evaluation have been used represent only a small fraction of the total (Morsh and Wilder, 1954), and of these, the variables measured have often been limited in breadth and peripheral in importance.

Methodological Problems

Criterion Problems

Definition of a Criterion. As in most other areas of research, one of the major problems in teacher effectiveness has been the definition of a criterion: What is good teaching? As Levin (1954) points out, this involves a value judgment; it is a question for which there is no research answer. "Science can tell us how to get there, but not where to go."(p.99). When a rater is asked to identify "good" and "poor" teaching, he is asked first for a value judgment, "What are the goals of education?", and second, he is asked, "Who is more effective in attaining these goals (based on the data of the classroom)?" It is the telescoping of these two processes which has led to so much confusion and conflicting results. There may be as many definitions of good teaching as there are raters, and this influences the choice of data gathered to answer the second question. As a consequence,

the consistencies which might have been found between classroom dimensions and pupil learnings have been obscured by differences in values.

Mitzel (1960) has analyzed the problem of the criterion of teacher effectiveness in very cogent terms,

> "...teacher effectiveness as a concept has no meaning apart from the criterion measures or operational definitions of success as a teacher. These measures should possess four basic attributes: (a) relevance, (b) reliability, (c) freedom from bias, and (d) practicality."
>
> "Relevance as a criterion attribute is the product of a rational analysis of the job functions and the job objectives. In so far as a criterion measure reflects the behaviors required in the achievement of job objectives, it is relevant. ...(But) analysis of tasks such as teaching does not readily yield criteria which are directly relevant to the job functions. As a consequence, we have often had to depend upon indirect and fragmentary definitions of teaching success. The difficulty stems, in part, from lack of agreement among educators on a hierarchy of goals and objectives for teaching." (p. 1482).

As other reviewers have also commented, this is essentially a value question, and not answerable by research. Mitzel continues,

> "Perhaps the most significant statement that can be made about any proposed criterion variable is that it is a partial criterion. Relevance for a particular criterion should be defended logically without undue apology for its lack of comprehensiveness, since it seems unlikely that we shall soon find a generally accepted formula for weighting school goals according to their importance for overall effectiveness. Relevance, whether direct or indirect is the paramount attribute of any criterion measure." (p. 1482).

A more promising approach would seem to be to abandon the attempt to get agreement on effective teaching, and within a conceptual framework, to identify and measure dimensions of the classroom which seem important; and to relate these to aspects of pupil learning and change--the ultimate criterion. The functional relationships between these variables can then be established empirically. Given consensus on some of the desired outcomes of education, these empirically established relationships can be used to

specify the characteristics of the classroom which should be optimal in fostering the valued kinds of pupil growth.

In this analysis of relationships, there is need for working within a conceptual framework, or theory. This sets two requirements of the research: that the variables selected for study are ones which can be related meaningfully to theory; and that the intervening phase of classroom behavior between teacher characteristics and pupil growth is specified. With respect to the first, numbers of studies have examined teacher characteristics which were available or convenient, such as marital status or age, which could scarcely be expected to contribute much to an understanding of the teaching process, even if significantly related to pupil learning. The second requirement has been well expressed by Levin (1954):

> "The most complete picture of the teaching process would be to predict from the antecedent conditions to the classroom behavior and in turn from behavior to pupil effects. Characteristically the middle step has been omitted, so that the result is a statement relating certain antecedent characteristics of the teacher to some aspects of pupil change. Although the procedure is legitimate, it presents only a partial picture. For example, there are some interesting findings that secondary school pupils taught by teachers who had neither a college major nor minor in the subject they were teaching scored <u>higher</u> on achievement tests than those students whose teacher had such college training. The findings are indeed perplexing. Notice that this research design has bypassed what the teacher is actually doing in the classroom. Without the added information we can only conjecture about these contrary-to-common-sense findings." (pp. 102-103).

While empirical findings of both sorts can be useful for selection and prediction, they do not help to extend or clarify educational theory or to modify practice or teacher education programs.

<u>Measurement Problems</u>. Various problems of measurement have contributed to the difficulty of research in teacher effectiveness, but the problems of ratings have been among the more serious and frequent. Among these are the

assumptions that the relevant aspects of the classroom can be observed, observed reliably, and given the appropriate weight in an over-all rating. It seems reasonable that some aspects of the classroom will be more accessible to the observer than others. Leiderman, Hilton, and Levin (1957) found the most predictable aspects of practice-teacher behavior to be amount of physical contact with pupils, amount of movement about the classroom, and to a much smaller degree the amount of support given the child and the amount of reward used by the student teacher. These appear to be some of the behaviors more accessible to observation, and to be predictable in the order of accessibility. When the rater is asked to combine different kinds of observations into an over-all rating, this assumes that he is able to weight these optimally. The bulk of the evidence on this issue supports the opposite conclusion that actuarial predictions are superior to "clinical" ones (Meehl, 1954).

Medley and Mitzel (1959) reviewed studies of the relations between ratings and other criterion measures of teacher effectiveness and carried out a study using more objective measures. Their review consistently found no relation between ratings of teacher effectiveness and pupil learning in the classroom. Their study compared ratings with more objective measures of the classroom than had previously been available, and with several measures of pupil learning. The only significant relationship found was that between the ratings and emotional climate in the classroom. On the basis of these results and their review, they call into serious question the use of ratings for any purpose other than as an indication of the emotional climate of the classroom.

Another difficulty has been the tendency for research to be done on very restricted aspects of the classroom or of pupil learning. This is a

problem of measurement--the unavailability of measures which adequately describe the criteria--as well as a problem of criterion complexity.

Criterion Complexity. A basic problem has been the tendency to select a measure or a few measures of teacher behavior or pupil change as "the" criterion.

1. Pupil Change Criteria. Among studies of pupil change, for example, subject-matter achievement has frequently been taken as "the" criterion. Results for these narrow aspects of the classroom are sometimes difficult to interpret even when significant differences are found. For example, one study (Brookover, 1945) found superior subject matter achievement on the part of pupils in the classrooms of teachers who provided less supportive and warm emotional climates. Less learning occurred in warmer, more supportive classrooms. A first question which occurs is whether the learning represented was learning of facts (or even overlearning), or whether it was a more understanding kind of achievement which would be more generalizable. But an even more important question is that of the extent to which other learnings were taking place in each set of classrooms. Very conceivably, pupils in the classrooms characterized by warm social-emotional climate were learning increased social skills and favorable attitudes toward the classroom in particular, and education and learning in general. If this were true, the evaluation of the outcome might be quite different; but the study does not provide data on this point and leaves the interpretation in doubt. In relation to this problem, Long (1957) comments,

> "Changes must include all-around pupil growth: intellectual, social, emotional, asthetic, and spiritual. The achievement of knowledge and skills is not necessarily a measure of understanding, interpretation, application, appreciation, or of reasoning. It certainly, in itself, is not an indication of personality growth, social adjustment, or other non-intellective functions. There has

> been considerable discussion about pupil change but few researchers, except in the area of achievement of skills and knowledge, have actually studied pupil change. That is why at this late date the following statement is found in the literature: 'There is some basis, then, for the undercurrent of feeling that researchers studying the problem of teacher effectiveness are no closer to the core of the problem than they were two decades ago.'" (p. 220).

2. Classroom Process Criteria. When teacher behavior is taken as a criterion, a problem related to but broader than the use of ratings is the treatment of teaching effectiveness as a unitary function. Even when an attempt has been made to rate separate dimensions, the results have often seemed to reflect "halo" to such a degree that only a single dimension was in fact being measured (Morsh and Wilder, 1954). In contrast to the frequent treatment of effectiveness as a unitary trait, several recent researches indicate the complexity of the basic dimensions that characterize classrooms. Medley and Mitzel (1958) have identified by factor analysis three dimensions of the classroom which are essentially unrelated, and Bowers and Soar (1961) have verified their essential independence. Ryans, in a comprehensive study of teacher behavior (1960), found three basic dimensions underlying both elementary and secondary teaching. While they were not independent, he felt they were unique enough that it was important to use them separately in further research. One of the reasons, which is relevant here, was that the separate dimensions were predictable by item analysis, whereas the composite was not.

Once a multidemensional view of the teaching-learning process is accepted, it seems likely that a teacher may do some things that are effective and some that are not; or that some acts may be effective in working toward one goal but not another; or even that a given act may be effective in working toward one goal, but inimical toward the achievement of another; or that an act may be effective in teaching one child and ineffective with another.

A Proposed Framework

A contemporary analysis of the criterion problem by Mitzel (1960) proposes a three category framework for conceptualizing criteria of teaching effectiveness. The first category is that of product criteria, which consists of changes in the pupil--presumably measureable aspects of growth, or change in attitude or behavior. He comments,

> "Considering the theoretical importance of product criteria in the assessment of teacher effectiveness, it is surprising that so few studies have used some measure of student growth as the operational definition of teacher competence. Barr's summary of 138 studies, published in 1948, lists only 19 different investigations which by any stretch of the imagination can be said to use a measure of student gain as a criterion. In 1956 Mitzel and Gross found only twenty studies which had used a student-growth criterion to measure teacher effectiveness in elementary or secondary schools." (p. 1483).

Second, he suggests the category of process criteria--aspects of classroom operation which are deemed to be worthwhile in their own right, although they may not be directly related to the outcomes of education, the product criteria. These process criteria are most often measurements of classroom climate or typical situations involving the social interaction of students and teachers. Other examples would be the extent to which teachers discipline students effectively, maintain rapport with students, or individualise instruction in accord with students' capacities and achievements. In relation to student behavior in the classroom, another process criterion, Mitzel says,

> "In considering both teacher behavior and student behavior as process criteria it becomes clear that neither of them should be studied in isolation from the other. The interaction between them appears to be the dominant aspect of the whole process of learning." (p. 1484).

The final criterion category suggested is called presage criteria--a label applied because it assumes predictive power on the part of the criteria. Some examples would be teacher personality, training, experience, intelligence, and social skills. Mitzel comments,

> "In a sense they are pseudo criteria, for their relevance depends upon an assumed or conjectured relationship to other criteria, either process or product. ...precedent forces their consideration as criteria, since the bulk of the research on teacher competence has employed dependent variables which fit into this category." (p. 1484).

It is clear that these categories have been emerging constructs in educational research for some time. Levin (1954), for example, wrote of the need to organize research around three steps: antecedent teacher characteristics, classroom behavior, and pupil change. It seems clear that these aspects of the developmental process are the ones to which Mitzel applies the labels presage, process, and product, respectively. Levin's delineation of the importance of examining and relating successive phases of the process and Mitzel's explicit specification of these phases should clarify the requirements of research in the future.

Development of Composites

When the several phases, and complexity within any one phase, of the possible criteria of teacher effectiveness are accepted, an important rational question appears to be that of whether prediction may be possible only by taking a more detailed view of the process and then synthesizing those resulting elements into a composite. As Ryans (1960b) says,

> "The predictability of teacher effectiveness undoubtedly is affected by the multidimensionality of the criterion. There is accumulating evidence that prediction can be accomplished with better than chance results for specified dimensions or components of the criterion. On the other hand, the prediction of over-all teacher effectiveness is possible only to the extent that some general agreement can be reached regarding the dimensions comprising over-all effectiveness (involving, of course, acceptance of a common set of educational values) and how they should be combined to form a composite." (p. 1490).

Short of the general agreement on over-all teacher effectiveness which is highly desirable but has not yet been attained, an alternative procedure

would be to accept a set of dimensions which are relevant by Mitzel's criteria and can be reliably measured, and examine the nature and utility of composites created by several statistical procedures. Very little research has been done on this aspect of teacher effectiveness, although Ryan's work (1960a) touches on the problem, and the Texas Talent Project (McGuire, Hindsman, King and Jennings, 1961) has synthesized pupil gain by a method which would be appropriate here.

Problems of General Methodology

Aside from the criterion problems already discussed, other methodological problems have made research and interpretation in this area difficult.

Choice of Subjects. Much past research has been done with student teachers or with recent graduates of teacher education programs. Typically, predictor data were collected during the training of the new teacher and criterion data were collected during her student teaching or her first year in the classroom. It has been generally accepted from informal observation that many teachers change their methods considerably during their early experience, so that they are likely to be quite different teachers even by the end of the first year than they were at graduation. This conclusion is supported by Morsh and Wilder (1954) and by Bach (1952), who found ratings of practice teachers to be unrelated to their first year inservice ratings. This may be a consequence of rater unreliability (Hampton, 1951), but whatever the cause, the finding would seem to limit generalization from student teachers to inservice teachers, and to raise the question of the stability of early inservice teaching.

Failure to Consider Pupil Differences. Much research on teacher effectiveness has treated pupils as though they were homogeneous, at the same time that it has conceded that this is not true. There is the ever-present probability that pupil differences may lead to opposite responses to the same teacher behavior, so that the evaluation of different teacher methods requires specification of the pupil characteristics in relation to which they are effective or ineffective. The work of Castaneda, Palermo, and McCandless (1956) suggests this. They found an interaction between anxiety level and complexity of learning, in which a higher level of anxiety was optimal for learning simple tasks than for complex. If children differ in their sensitivity to induction of stress by subject-matter problems, peer group pressures, or teacher behavior (which seems inescapable), then the same teacher behavior which leads to change in tension level in the classroom would be expected to facilitate learning for one pupil while handicapping it for another. Similarly, interactions between pupil ability level and the relative effectiveness of conceptual or drill teaching have been reported.

Summary

In summary, when all of these various aspects of the teaching-learning process are considered, it seems likely that effective research will require the identification and measurement of a variety of characteristics of the classroom, the simultaneous measurement of a number of dimensions of pupil learning, and the control or measurement of a number of pupil characteristics. Then it may be possible to find characteristics of the teacher which predict a classroom in which effective learning occurs, and to improve teacher education on the basis of a better understanding of the teaching-

learning process. Perhaps it will also be possible ultimately to relate aspects of pupil growth to the effectiveness, happiness, and responsibility of the mature adult to which this educational process contributes.

Research Results

Correlates of Teacher Effectiveness

Results of research on the correlates of teacher effectiveness will not be reviewed in detail here. The methodological problems cited earlier that have plagued past researches have limited their meaning, while other summaries available or in press are so comprehensive as to make a review here of little value. Barr (1948) summarized and integrated the literature to that time, and Morsh and Wilder (1954) have summarized later studies and presented an integration that is both intensive and extensive. Other less extensive integrations are available in Ackerman (1954), Fulkerson (1954), Long (1957) and the American Association of School Administrators (1961). Bibliographic summaries are available in Barr (1955, 1958), Tiedeman and Cogan (1958), and Watters (1954). Gage (in press) is expected to present a current and extensive integration of the work to date.

The most concise of the recent integrations of the research on correlates of teacher effectiveness is that of Ryans (1960b). He summarizes as follows:

> "Measured intellectual abilities, achievement in college course, general cultural and special subject-matter knowledge, professional information, student teaching marks, emotional adjustment, attitudes favorable to students, generosity in appraisals of the behavior and motives of other persons, strong interest in reading and literary matters, interest in music and painting, participation in social and community affairs, early experiences in caring for children and teaching (such as reading to children, taking a class for the teacher), history of teaching in the family, size of school and size of community

in which teaching, cultural level of community, and participation in avocational activities, all appear to be characteristics of the teacher which are likely to be positively correlated or associated with teacher effectiveness in the abstract.

"Extensiveness of general and/or professional education, enrollment in particular professional courses, personal appearance, and grade or subject taught (with some exceptions) appear to bear very little relation to the abstracted criterion.

"Age of the teacher and amount of teaching experience seem to manifest an over-all negative relationship with teaching effectiveness, although there is evidence of curvilinearity, increase in effectiveness being positively correlated with experience during the early years of teaching careers.

"Sex differences in teacher effectiveness do not appear to be pronounced among elementary teachers, although at the secondary level it appears that women as a group may be more effective than men as a group on specified criterion dimensions.

"For teachers of all levels considered together, there seems to be little difference in specified dimensions of teacher behavior between single and married teachers; within the elementary school the evidence somewhat favors married teachers; at the secondary level unmarried teachers as a group appear to be superior with respect to criterion measures which relate to systematic, responsible teacher classroom behavior." (p. 1490).

Factor Analyses of Teacher Behavior

Several studies have presented factor analyses of data in an attempt to identify basic dimensions of teacher behavior, teacher-pupil behavior, or teacher personality and behavior. The need for analysis and synthesis to reduce the number of traits believed to be important is evident. Often, however, the analyses have been carried out with the teacher behavior described by ratings, so that some of the same difficulties of subjectivity of measurement cited earlier limit the interpretation of these studies.

Since no other basis for organization appears more useful, these studies will be summarized chronologically, in the hope that methodology has improved over the years. Only studies which included data on teaching competence or behavior are summarized.

Hellfritzsch (1945) factor analyzed data for 24 teachers which included a number of aspects of teacher ability based on tests of intelligence, subject matter knowledge, and general information, as well as attitude scales, measures of pupil learning, and a variety of ratings of teacher performance. The factors which emerged were:

I. General Knowledge and Mental Ability;

II. Teacher Rating Scale, related to supervisory ratings of the teachers, apparently reflecting the rater's liking for the teacher and little else;

III. Personal, Emotional, and Social Adjustment;

IV. Eulogizing Attitude Toward the Teaching Profession.

The study was replicated with 72 teachers and the same four general factors were found in the same order. A fifth factor, Teaching Ability, appeared this time which reflected pupil gain, but Hellfritzsch did not label it a general factor because of the few measures which loaded on it.

Schmid (1950) analyzed a number of measures of personality and attitude, along with grades both for education courses and other college work for 102 education students, equally divided between male and female. The results of an oblique analysis for the females indicated four common factors:

I. Problems Response Set, an awareness of problems actually felt by the student;

II. Professional Maturity, made up of "youthfulness, undifferentiated interest pattern, poor professional educational adjustment, and low grades on education courses." (p. 314);

III. Introversion;

IV. Social Adjustment.

From the analysis of the data of the male students, only two factors were obtained:

I. Social-educational Adjustment, made up of "purpose" and general social adjustment;

II. Personality-psychological Factor, made up of a series of personality scales whose meanings were not analyzed.

Lamke (1951) factor analyzed the responses of 18 teachers (10 rated good and 8 rated poor by university supervisory staff and principals) to Cattell's 16 P.F. Test. He concluded that good and poor teachers did not fall into two well-defined patterns but that several patterns existed for each group. He described the differences as follows:

> "Using Cattell's terminology, it appears that good teachers are likely, more than poor teachers, to be gregarious, adventurous, frivolous, to have abundant emotional responses, strong artistic or sentimental interests, to be interested in the opposite sex, to be polished, fastidious and cool. Poor teachers are more likely than good teachers to be shy, cautious, conscientious, to lack emotional responses and artistic or sentimental interests, to have comparatively slight interest in the opposite sex, to be clumsy, easily pleased, and more attentive to people." (p. 253).

It would seem appropriate to question how many differential patterns of behavior can be identified on a sample of 18 people, however; and it seems likely that differences in rater values may account for at least some of the differences in patterns found among "good" and "poor" teachers.

Bach (1952) analyzed college records, ratings of performances in practice teaching, and ratings in first year inservice teaching for 76 students. Four oblique factors were found to account for the intercorrelations of the final practice teaching ratings:

I. Pupil Response, primarily to the learning experiences presented by the teacher;

II. Technical Competence, subject matter and professional understandings (which correlated .61 with Factor I);

III. Relations with Others, (which correlated .31 with Factor I);

IV. Personal Appeal, including appearance, (which correlated .45 with Factor III).

The results from the analysis of beginning teacher data for the same subjects resulted in four oblique factors.

I. Technical Competence (as Factor II above);

II. Cooperative Attitude, personal and social adjustment;

III. Initiative, including skill in presenting learning experience;

IV. Personal Appeal, as Factor IV above.

The only significant correlation in this analysis was .60 between Factors I and II.

Medley and Mitzel (1958), in the development of the Observation Schedule and Record, collected extensive observational data on 49 teachers in such a form that the observer's task was limited to recording the occurrence or non-occurrence of specific bits of behavior. Scoring and combining data were done separately. These different kinds of observations were grouped rationally into clusters that seemed to "belong" together, and the clusters factor analyzed. Three orthogonal factors emerged:

I. Emotional Climate, in which teacher hostility and pupil disorder were negatively weighted, and teacher supportive behavior was positively weighted;

II. Verbal Emphasis, with time spent on reading and verbal activities positively weighted, and time spent on social studies negatively weighted;

III. Social Structure, with autonomous social groupings positively weighted, and teacher verbal output negatively weighted.

Reliabilities ranged from .77 to .90. A more detailed description of procedure, instrument and results will be found in Chapter III.

Solomon (1962) studied a variety of categories of teacher behavior ranging from relatively gross subjective ratings of classroom procedure such as amount of direct control and relaxation of atmosphere, to more minute aspects of teacher behavior, derived from categorizing units of speech from tape recordings. Altogether, 169 variables were available, from 24 college teachers of evening credit courses and their classes. Factor analysis and varimax rotation produced eight factors:

I. Permissiveness vs. Control;

II. Lethargy vs. Energy;

III. Aggressiveness vs. Protectiveness;

IV. Incoherence vs. Coherence;

V. Emphasis on Content vs. Emphasis on Student Growth;

VI. Dryness vs. Flamboyance;

VII. Emphasis on Student Participation vs. Emphasis on Lecture;

VIII. Warmth vs. Coldness.

Criterion data against which teacher scores on these eight factors were related consisted of a pre- and post- factual test and comprehension test administered to the students, resulting in gain scores; and student evaluations of interest in the course and of teacher effectiveness.

When teacher standings on each of the eight factors were related to mean class scores on the four criterion variables, the following relations were found: for Factors I, III, V, and VII, no significant relationships;

for Factor II, comprehension gain was associated with instructor energy; for Factor IV, instructor coherence was related to student interest, student evaluation of the instructor, and factual gain; for Factor VI, instructor flamboyance was related to student comprehension gain; for Factor VIII, instructor warmth was related to student evaluation of the instructor.

The work which is both recent and performed on the largest number of subjects is that of Ryans (1960a). Ratings of 275 elementary and 249 secondary teachers were collected on 25 seven-point rating scales, six of which reflected pupil behavior. When these were factor analyzed, orthogonal rotations failed to produce approximations of simple structure, so oblique rotations were performed. Although somewhat different factor structures were found for the two groups, three factors were clearly common to the two groups. They were designated as:

> "TCS Pattern X_o, reflecting understanding, friendliness, and responsiveness vs. aloofness and egocentrism on the part of the teacher; TCS Pattern Y_o, reflecting responsible, businesslike systematic, vs. evading, unplanned, slipshod teacher behavior; and TCS Pattern Z_o, reflecting stimulating imaginative, original, vs. dull, routine, teacher behavior." (p. 102).

Further study of these patterns was carried out on over 1500 elementary and 1900 secondary teachers, altogether. The results replicated, and the patterns were found to be relatively highly correlated for elementary teachers, but less highly related for secondary teachers.

If an attempt is made to group similar factors from different studies, with willingness to tolerate some looseness and ambiguity, beginnings in grouping can be made, but many findings still stand alone. One of the more general findings of studies in which personality data were included is a factor of personal-social adjustment (Hellfritzsch's Personal-Social,

Schmid's Introversion and Social Adjustment, and Bach's Relations with Others). A general classroom management factor (including the materials of learning) seems to emerge from several (Bach's Technical Competence and Pupil Response, Ryans's Systematic vs. Unplanned, and perhaps Medley and Mitzel's Verbal Emphasis and Social Structure); as well as a Personally Interesting and Attractive factor (Hellfritzsch's Rating Scale, Bach's Personal Appeal, Ryans's Stimulating vs. Dull, and Solomon's Dryness vs. Flamboyance, and perhaps his Lethargy vs. Energy). Accepting Teacher Behavior is found in three studies (Ryans's Understanding vs. Aloofness, Medley and Mitzel's Emotional Climate, and Solomon's Warmth vs. Coolness). The remainder of the factors appear to scatter.

There seems to be a similarity between the accepting behavior and classroom management dimensions and the two dimensions of "Consideration" and "Initiating Structure" found in the Ohio studies to summarize much leadership behavior (Fleishman, 1953). This seems a reasonable finding, since the teacher is the classroom leader as well as a subject matter resource.

Ryans has summarized the importance of information such as these studies provide. He comments (1960a):

> ". . . the problem of how the personal and social characteristics manifested in the classroom behavior of the teacher are organized is important from both the theoretical and practical viewpoints. If education is going to be at all concerned with the components of teacher behavior, it is imperative that something be learned of how these components are organized; or, if there is no discernable organization among them, this should be known. The problem cannot be ignored or begged." (p. 96).

Although numbers of factor analyses have been carried out toward understanding the organization of teacher behavior better, none have applied the procedure to a simultaneous examination of teacher behavior based on observation rather than ratings, and of teacher personality.

Summary

When the literature of research on teacher effectiveness is viewed as a whole, the promising advances appear to be in the understanding of the problems of doing research in a difficult area, and a beginning (The Observation Schedule and Record) toward dealing with one of the major problems, objectively recording and quantifying the behavior of teachers and pupils in the classroom. The hope of future progress appears to lie in the objectivity and multidimensionality of newly developed measures in this area, and in the complex statistical procedures which will make possible the simultaneous utilization of these several dimensions.

CHAPTER III

PROCEDURE

This chapter will present briefly the objectives of the study, will describe the subjects, measures, and procedures employed and outline the statistical analyses completed on the data.

Objectives

Considering the history of past studies which have attempted to predict teacher effectiveness, a new study in this area would appear to require justification on the basis of bringing something new or different. Beginning with a multivariate view of the teaching-learning process, this study is an integration of newly-developed measures of classroom behavior and complex statistical procedures, in order to study the nature of the composites which are created, the relative utility of these composites, and the dimensions basic both to these composites and antecedent teacher characteristics.

Since this was not intended to be an experimental study, but rather an exploratory, empirical one, the goals of the research are stated in terms of objectives rather than hypotheses:

1. To create composites of teaching effectiveness through the use of two multivariate statistical procedures: (a) canonical correlation, and (b) scaling by reciprocal averages, and to compare them with an average of standard scores.
2. To study the relative predictability of teaching effectiveness by canonical correlation, and by multiple correlation.

3. To seek the basic dimensions underlying both teacher-pupil classroom behavior and teacher personality and attitudes by factor analysis.
4. To test the utility of predicting a composite of teacher-effectiveness by item analysis of personality measures.
5. To cross-validate the results of these analyses.

Subjects

In view of the low and inconsistent relationships found in previous studies between student teacher classroom behavior and the behavior of the same subjects as inservice teachers, it seemed desirable to draw on inservice teachers as subjects. This was done for both samples of data.

The First Sample

A detailed description of the first sample of teachers is contained in the project report (Bowers and Soar, 1961), and summarized in Appendix A in this report. Briefly, they were a group of fifty-four elementary school teachers from the city of Nashville, Tennessee, and the surrounding suburban areas. They were recruited by holding meetings initially with principals of the schools, then with teachers who might be interested from a given school, and often finally an individual interview was held with each of the teachers who expressed interest. This was accompanied by a full description of the nature of the project, the time commitment which would be required of the participating teacher, and the aims and goals of the project. The sample which was obtained by these means was one in which typically one or two teachers were recruited from a given school. Numbers of schools were contacted from which no teacher volunteered, and from a scattering of schools two or more teachers volunteered.

The greatest number obtained from any one school was seven, and this was a school in a high socio-economic neighborhood which was well known for the morale of its staff. The schools sampled were distributed from the lowest to the highest socio-economic areas, but all were in relatively well-paying districts.

Inevitably this was a biased sample, one made up of teachers who were particularly interested in the welfare of school children and in increasing their own teaching skills. Many of them were mature teachers with extensive experience; several were close to retirement. There seemed very little question but that this was a select group of teachers.

The Second Sample

The teachers in the second group were a rather different sample. They were obtained from the urban-suburban Columbia and Sumter areas in central South Carolina; five entire school staffs comprised the bulk of the sample, with eight additional teachers from two other schools. Recruiting was carried out by meeting first with principals of a number of schools, then with the staffs of those schools which the principals felt might be interested. In all but the two schools mentioned, staffs as a whole agreed or declined to participate. So far as is known, no administrative pressure was applied to any of these school groups to encourage participation; and indeed, there was mutual agreement between the investigator and administrators of the schools that this would be the case. Ninety elementary teachers were recruited originally, and classroom measures were collected for all of these. However, two failed to complete the personality and attitude measures satisfactorily and had to be dropped from the study, leaving a sample of 88.

The socio-economic areas in which the schools were located ranged from high in the case of one school to a mill district, low socio-economic area for another school, with the others distributed in between. All were from comparatively well-paying districts. To the extent that urban-suburban, well-paying districts may produce a bias, these teachers are probably a select group. Otherwise, a minimum of bias may have occurred, at least as they represent teachers in South Carolina.

Comparison of Samples

It seems likely that a broader representation of teacher ability, personality, and attitude should be represented in the second sample, since entire school staffs were used, as contrasted to the kind of selection that occurred in the first sample in which volunteer teachers from scattered schools were used. In the first study, presumably only the teachers who felt most secure in being observed participated; whereas it seems likely that at least some of those in the second sample would not have volunteered had it been an individual affair but were not willing to refuse to participate when the decision was made on a school-wide basis. When all of these factors are considered, it seems probable that the second sample was a less select, more representative one than the first.

All teachers in both samples were white elementary teachers, most of them married, few without training at least at the B.A. level, and some with advanced training of varying degrees. A detailed summary of the description of both samples will be found in Appendix A, along with the comparisons of the South Carolina sample with teachers of the entire state.

Additional data will be found in Table 19, Chapter IV, in which the means and standard deviations of the attitude and personality characteristics of both samples are compared. Significant differences were found in the Survey measures and two personality measures. The former may be large enough to be practically important, but this seems unlikely for the latter two. The implications of the differences are discussed in Chapter IV on page 86 ff.

Measures

The measures employed included inventory measures of attitude and personality of the teacher, an observational measure of three dimensions of teacher-pupil classroom behavior, and a situational test of pupil group problem solving skill.

Bowers Teacher Opinion Inventory (BTOI)

The test author has described the inventory as follows (Bowers and Soar, 1961):

> "The BTOI provides a measure of satisfaction with teaching that is relatively independent of situational and environmental factors. It measures those attitudes that condition the relationship between the teacher and the profession. Criteria for the satisfied teacher included five factors: (1) continued teaching in spite of greater financial reward from other occupations, (2) social services to the community, (3) specialized professional training, (4) active participation in professional organizations, and (5) improvement in professional techniques. In the initial validation study, the BTOI did not correlate significantly with the MTAI. Additionally, concurrent validity was studied in terms of principals' ratings and found to compare favorably with the concurrent validity of other attitude scales. Reliability was estimated to be .958. The research on this inventory seemed to indicate its usefulness in assessing those professional attitudes that predispose a person to be a permanent member of the teaching profession." (p. 29).

Minnesota Teacher Attitude Inventory (MTAI)

This is an inventory designed to measure teacher attitudes which are expected to be related to the kind of teacher-pupil relationships that will be found in her classroom. The test authors, (Cook, Leeds, and Callis, 1951), describe the classroom atmosphere expected of the high scoring teacher as follows:

> "The teacher should like the children and enjoy teaching. Situations requiring disciplinary action should rarely occur. The teacher and pupils should work together in a social atmosphere of cooperative endeavor, of intense interest in the work of the day, and with a feeling of security growing from a permissive atmosphere of freedom to think, act and speak one's mind with mutual respect for the feelings, rights and abilities of others. . . . Group solidarity resulting from common goals, common understandings, common efforts, common difficulties, and common achievements should characterize the class.
>
> "At the other extreme of the scale is the teacher who attempts to dominate the classroom. He may be successful and rule with an iron hand, creating an atmosphere of tension, fear, and submission; or he may be unsuccessful and become nervous, fearful and distraught in a classroom characterized by frustration, restlessness, inattention, lack of respect, and numerous disciplinary problems." (p. 3).

In the development of the inventory, items were written which were expected to tap attitudes underlying this dimension and were validated by their ability to discriminate between teachers identified by principals as having unusually good relationships with pupils and those identified as having poor relationships with pupils. The final form of the inventory was made up of items which had survived rigorous screening and which in cross-validation showed validities mostly in the upper fifties against ratings of teachers by principals and "experts."

Reliabilities above +.90 have frequently been found.

Other findings supporting the validity of the MTAI have been reported by Callis (1953), Della Piana and Gage (1955), Kearney and Rocchio (1955), Hoyt and Cook (1959), and Coleman and Collett (1959). A failure to support

its validity has been reported by Rabinowitz and Rosenbaum (1958), in which MTAI scores of teachers correlated -.01 with pupil responses on the My Class inventory.

Generally, however, the MTAI has been found to be related to ratings, by pupils or administrators, of the warmth and supportiveness of the emotional climate created by a teacher. There is little question but that this is the most used inventory for the purpose at the present time.

The Survey of Educational Leadership Practices (Survey)

This is a 50 item questionnaire which places the teacher's ideal of classroom leadership on a democratic-autocratic continuum. The author (Valenti, 1952), describes this dimension as follows:

> "At one extreme the teacher defines his role in terms of impersonal, direct, formal, and one-way communications in the interaction process in order to maintain his own and organizational status, and depends upon a 'technical' approach. At the other extreme, he uses personal, indirect, informal, and two-way communications in order to develop the pupil and the class and depends upon a 'human' approach." (p. 37).

Although the inventory has not been widely studied, a validity coefficient of .59 was found in the original development between inventory scores and ratings by colleagues and supervisors. Reliabilities ranging from .61 to .88 were obtained for sub-scores, with reliability for the total test unreported.

In addition to the description of an "ideal" which the teacher expresses, a second set of responses are recorded which are descriptive of the "actual" leadership role the teacher feels she plays, and a "discrepancy" score was calculated in this study as a measure which might be as meaningful as either of the original scores.

The Minnesota Multiphasic Personality Inventory (MMPI)

The validity keys and the ten clinical keys were used in the analyses of both samples of data, as well as several special keys which seemed to be particularly relevant to the work of the teacher. Since the MMPI is probably the best known and most widely used of the structured personality inventories, the dimensions measured by the clinical keys will not be reviewed here. Where keys predicted aspects of teacher effectiveness, the probable interpretation of the keys will be discussed in that context in terms of the relevant description of personality; otherwise information on the inventory may be found in Welsh and Dahlstrom (1956) and in Dahlstrom and Welsh (1960). There is very little question that the MMPI is the structured inventory whose empirical validity has been most thoroughly studied.

Special Keys. The special keys that were also scored were: Welsh's A and R, Barron's ES and the Cook-Medley Ho, FV, and TA scales. The descriptions of these scales are as follows:

1. Anxiety (A). This scale was developed by Welsh to represent the first factor which has usually been found in factor analyses of the MMPI. Although named Anxiety, the content of the scale reflects slowness in thinking, negative emotional tone, lack of energy, pessimism, and personal sensitivity.

2. Repression (R). This measures the dimension which usually emerges as the second factor in analyses of the MMPI. It reflects concern about health, physical symptoms, emotional adjustment, stimulation by social situations, unaggressiveness, and lack of social dominance. High scorers on the scale are characterized by repression and denial; low scorers, by externalized and "acting-out" behavior.

The A and R scales were included in the study because Welsh found the two scales together summarized much of the reliable variance of the MMPI (Welsh and Dahlstrom, 1956, pp. 264-8). As will be noted in the section on factor analytic studies of the MMPI, this is representative of the school of thought that two dimensions account for most of the individual differences which the inventory as a whole assesses.

3. Ego Strength (ES). This scale was originally built as a predictor of success in psychotherapy, but collateral study with it has indicated that it may offer promise in situations in which general personality integration, adaptability, and resourcefulness are important. It correlates moderately with intelligence, general energy level, self-confidence, breadth of interest, tolerance and lack of ethnic prejudice, and social ease (Welsh and Dahlstrom, 1956; pp. 226-34).

4. Hostility, Pharisaic Virtue, and Teacher Attitude (Ho, PV and TA). These keys were developed by item analyzing MMPI records of extreme criterion groups selected on the basis of the MTAI (Cook and Medley, 1954). Further selection of items took place on the basis of content to make up the first two scales; Teacher Attitude is the sum of Ho and PV. A person scoring high on the Hostility scale is one who has ". . . little confidence in his fellow man. He sees people as dishonest, unsocial, immoral, ugly, and mean, and believes they should be made to suffer for their sins." (pp. 417-18). High scores on the PV key suggest preoccupation with ideas of sin and punishment and rigid moral values. Reliabilities for the keys range between .85 and .90 and correlations with the MTAI are generally in the upper forties.

Research Relating the MMPI to Teacher Effectiveness. The MMPI has been related to teacher effectiveness in a number of studies, but most of these used ratings as the criterion measures and student teachers as subjects. The problems raised by these procedures have been discussed in Chapter II, but a number of the studies will be reviewed briefly. A more extensive review has been presented by Rippy (1960).

Medley and Williams (1957), (not using student teachers or ratings) found that Ho and PV predicted pupil liking of the teacher (as measured by the My Class inventory) significantly. Neither inventory predicted growth in reading, however. Michaelis (1954) found that Pt and Sc differentiated ratings of student teachers, but not at the one per cent level. Tyler (1954) found little relation between practice teacher ratings and MMPI scores, but questioned the reliability of his criteria. Tanner (1954) attempted differentiation of practice teachers rated "good" and "poor", and found that only K distinguished the two men's groups at the five per cent level. The women's groups were distinguished by K, D, and Pt. Gowan and Gowan (1955) developed a key for teacher prognosis (TP) by item analysis which appeared to be successful in predicting ratings. Moore and Cole (1957) found the sum of T scores on the clinical keys to be related to the practice teaching ratings of a restricted group of students on whom raters agreed. The TP key was not significantly related. Michaelis and Tyler (1951) found Hy to be related to practice teaching ratings at the five per cent level, with Pd and Pt approaching significance.

Gough and Pemberton (1956) developed a configural scoring procedure for predicting practice teacher ratings, and found eight "signs" with some discriminating value on cross-validation.

It is interesting to note that Pt has most often emerged as the significant predictor, with K, D, Hy, Pd and Sc also appearing. It is not at all clear to this reviewer that these particular keys would be the ones which theory would suggest as the effective discriminators. After the fact, however, some of them can readily be integrated into theory (see Chapter IV, results of the canonical correlation).

Factor Analytic Studies of the MMPI. There appear to be two points of view about the factorial structure of the MMPI (Lingoes, 1960) which differ in emphasis. One is essentially a two-factor approach; Wiener (1948) used the categories of "subtle" and "obvious," and Wheeler, Little, and Lehner (1951), "neurotic" and "psychotic." On the other hand, there is the point of view supported by Lingoes (1960) and Harris and Lingoes (1955), which offers considerable support for a minimum of seven factors and the possibility of more.

The difference in these points of view is at least partially one of emphasis. Messick and Jackson (1961) support the two-factor theory, but comment that ". . . only two major factors and two or three minor ones are necessary to account for interrelations among the scales." (p. 300). They reviewed the literature in relation to this question, and correlated the loadings on the first factor (usually a "psychotic" factor) with the proportion of items keyed "true" in each scale. With few exceptions they found high correlations, four of eleven exceeding .85. From this they concluded that acquiescence is a major factor underlying response to the MMPI. They reviewed research by others indicating that social desirability is a likely second major source of variance. They suggest that much of the variability of the different scales of the MMPI might then be accounted for by differential weightings of these two stylistic response tendencies.

The opposing point of view, that the MMPI is considerably more complex than a two-factor interpretation would suggest, is supported by Lingoes (1960). He carried out a series of independent factor analyses based on the responses of four groups of subjects to a large number of keys for the MMPI. The scales that were studied included 28 of the Harris and Lingoes subscales (earlier derived by content analysis of the clinical keys), and 8 of Wiener's scales. The four groups included 150 male and 100 female neuropsychiatric patients, and 100 male and 100 female community subjects. The responses of each group were factor-analyzed separately, and rotated to the varimax criterion. Reproducibility of the factors was then investigated by use of the Burt-Tucker index of factor similarity. Seven factors were reproduced with an index of factor similarity of at least .70 in all four analyses, four additional factors were replicated in three analyses, and three further factors appeared in two analyses. The seven factors identified were:

I. General Maladjustment, reflecting generalized anxiety, depression, guilt, pessimism, and anxious worrying.

II. Denial of Social Anxiety, reflecting "social ease, interpersonal dominance, extraversion, and social adaptability (middle scorers) and veiled agression, social obtuseness, denial of social anxiety, and naïveté (high scorers)." (p. 79).

III. Loss of Control, reflecting ". . . at the high end, the failure of neurotic defenses, and in the middle ranges, inadequate control mechanisms." (p. 79).

IV. Denial of Distrust and Hostility, reflecting ". . . naïveté, denial of distrust and hostility, dependence upon the good will of others, impunitiveness, tolerance, repressive and submissive

behavior. . . . whereas (Factor II) deals with interaction case, the present factor was more specifically concerned with defenses against both hostile and aggressive impulses." (p. 80).

V. Familial Discord, reflecting ". . . intrafamilial conflict, unresolved dependency strivings, feelings of being mistrusted, misunderstood, deprived, and isolated." (p. 80).

VI. Inhibition and Apathy, reflecting ". . . inhibition, overcontrol, rigidity, obstinacy, feelings of apathy and inertia, emotional constriction, etc." (p. 80).

VII. Social Nonconformity, reflecting ". . . rebelliousness, argumentativeness, immaturity, hostility directed specifically against authority, potential for acting out. . . ." (p. 80).

In summary of these findings Lingoes comments:

"Admittedly, the number of factors obtained is partly a function of the number of variables included and may be to some extent an artifact resulting from item overlap. Nevertheless, to conclude that because the first two or three factors account for roughly half the variance that the test is just measuring this or that aspect of personality, e.g., sickness or general maladjustment, or is influenced primarily by some response set, is not justified on the present evidence." (p. 81).

The Observation Schedule and Record (OSCaR)

In the rationale of the development of this recording schedule, Medley and Mitzel (1958) comment on the use of ratings as criteria of teacher effectiveness, and in this context say:

"Whatever value such ratings have arises from the fact that they are based on observations of the teacher while he is teaching; the most serious of the limitations inherent in them arise from the fact that the evaluative judgment of the rater intervenes between the behavior and the score supposed to reflect it. There are at least two sources of variation introduced here that can only attenuate the validity of the ratings as measures of behavior. The cues upon which the observer bases his judgment and the relative weights assigned to them are both allowed to vary from observer to observer to some unknown degree, and this variation distorts the measured differences between teachers. . . .

> "By providing a schedule for recording behavior listing the cues to be responded to, the first source of error may be virtually eliminated. By making the assignment of weights a clerical task done by someone other than the observer, the second may also be made negligible." (p. 86).

The goal of the OSCaR, then, was to reduce the task of the observer insofar as possible to that of recording the occurrence or non-occurrence of specific classroom behaviors, or to tallying the frequency with which they occurred. Ideally, the behaviors should be simple and objective enough so that there would be little or no question as to whether or not the occurrence had taken place. The development of the schedule began with the gathering of a wide variety of specific bits of behavior that might or might not occur in the classroom -- specific teacher-pupil interactions, the kinds of materials being used, the kinds of pupil groupings observed in the classroom, and tallies of each teacher statement on a continuum of emotional warmth. In the development of the schedule, teachers were observed in their classrooms. The analysis of results followed several steps: first, items were selected which had exhibited reliable differences from classroom to classroom, then were combined rationally into clusters of items that appeared to "belong" together. These clusters were then factor analyzed, and the records of the various observations were scored on the resulting factor dimensions to estimate the reliability of teacher behavior on these dimensions. Eight of the fourteen clusters were found to load to a sufficient degree to warrant placement on one of three orthogonal axes. The first factor was Emotional Climate, on which manifest teacher hostility and disorderly pupil behavior (both negatively weighted) and supportive teacher behavior loaded; the second was Verbal Emphasis, on which time spent on reading and verbal activities were positively weighted and time spent on social studies negatively weighted; the third factor was Social Structure, for which autonomous

social groupings was positively weighted and teachers' verbal output negatively weighted. Reliabilities ranged from .77 for Verbal Emphasis to .90 for Emotional Climate.

The time required for the observation is slightly more than half an hour in the classroom, and the observer does no scoring or evaluation--he simply records. Later, the subkeys for each factor are converted to standard scores, and combined with equal weights and the indicated signs and this total score converted to a final standard score for each factor. A copy of the observer's schedule used in this project and the scoring procedure for data reduction are in Appendix B.

Russell Sage Social Relations Test (RSSR)

This is a test of the ability of an elementary school group to plan and carry out a cooperative group task (Damrin, 1959). It is structured in a way that makes it very difficult for a few individuals to assume responsibility for the group in the completion of the task, or for individuals to solve the problem without some kind of cooperative agreement. It is administered by a two-person team who assume control of the classroom. One person serves as administrator and presents materials and leads the pupil planning, while the other serves as observer and categorizes on a standardized observation schedule numbers of aspects of the activity of the pupil group, including each statement made in the planning session. The materials that are used are 36 interlocking plastic blocks in three colors and two shapes, out of which three different figures are to be constructed. The general procedure is one in which the observer (and the teacher, if she remains) sits at the back of the room, while the administrator works at a table or desk (cleared) at the front of the room. The administrator initially instructs the pupils that this is a test in which they are to work together

as a group and will receive a score as a group, then shows them the first figure and asks them what it is. When the pupils have identified it, they are then encouraged to plan how they are to build it out of the materials which are distributed throughout the class. The administrator role in the planning session is specified in the same manner as that of the examiner in an individual intelligence test, although not in as completely standardized fashion. Essentially, the role is one of laissez-faire leadership, in which responsibility for the direction and control of the planning session is left up to the pupil group. The administrator will provide the functions of maintaining order, asking questions to facilitate planning, or to encourage participation, but only under specified conditions, and with these actions, in effect, scored against the group. When the group has developed a plan with which it is satisfied, the administrator steps aside and begins timing the construction phase (the planning phase has no time limit, and this is made clear to the pupils).

In the construction phase pupils face, sometimes for the first time under a particular teacher, a situation in which there is no adult control provided. The behaviors exhibited presumably reflect the degree of responsibility pupils have learned to assume for their own behavior, and range from matter-of-fact, business-like progress with the task at hand, to wild, unrestrained running around the room, bickering and fighting. The materials in the task provide a highly motivating situation even for upper elementary grades, so that excitement usually runs high and a major part of the problem is that of responsible self-control in the group.

When the second problem is presented, the group has the opportunity to profit from the experience of the first problem. There are such possibilities as delegating the task to pupils with relevant skills, building

portions of the problem at different locations so as to minimise crowding around the construction task and speed the building, and planning what pupils should do after completing their part of the construction task in order to minimise interference with those who are still working.

In the planning stage, the observer records the proportion of pupils who appear to be involved in the activity at each of several stages, the nature of the plans proposed, whether the plans build on earlier plans, and the extent to which administrator leadership is required by the group. In the operations stage there is systematic recording of the proportion of involved pupils, the social-emotional climate of the work group, and the kinds of activities carried out by pupils who have left the construction task.

The test is scored on ten dimensions of the group's activities following general rules of procedure outlined in the manual. However, in both the first and second samples, the results of these ten dimensions were pooled to provide an over-all estimate of the performance of the group using empirical weights developed by Medley and Mitzel by scaling by reciprocal averages (private communications).

In the first sample, the reliability of this pooled score was estimated (Hoyt and Stunkard, 1952) to be .82; the intercorrelations between problems were about .90. The total score correlated with grade level approximately .70, consequently this effect was extracted statistically as indicated in the section on data collection and reduction.

In the first sample of data, numbers of classes failed to complete (or even reach) the second problem, and consequently never reached the third construction task; since the inter-problem correlations were high, the third problem was not administered in the second sample of data.

Data Collection and Reduction

Data collection followed the same pattern for both samples. At the same time a teacher volunteered to participate, or shortly after, she was given an envelope which contained a coded face sheet with spaces for her name, grade, and school. The envelope also contained the four inventories with their answer sheets, and instructions for the completion of each, in addition to the standard instructions which were printed in the question booklets. These were left with the teacher for completion on her own time and at her convenience. Since recruiting was carried out during the greater part of the spring semester, the time set as a deadline for completion of the inventories was two weeks after the end of school. The only identifying data on any of the answer sheets, however, was the code number which appeared on the face sheet. This procedure was followed to preserve the anonymity of the respondent in all of the data processing which involved the answer sheets themselves. Standard scoring procedures were followed.

For both samples, administration of the Russell Sage Social Relations Test (RSSR) was scheduled and completed for each of the classrooms, and observers made three unscheduled visits to each classroom to complete the observations with the Observation Schedule and Record (OSCaR). Two data collection teams were used; one which administered the RSSR, and another which completed OSCaR observations. Both members of the RSSR team learned both administrator and observer roles, and alternated. Scoring was by consensus of the two team members. The OSCaR observers were trained together and observed together on a regular schedule, comparing observations in order to assure continuing reliability.

As noted earlier, numbers of the lower grade classrooms in the first project were unable to complete the second problem of the RSSR, and consequently were not administered the third. A few lower grade classrooms were unable to complete problem one, and consequently were not administered problem two. In order to provide problem two scores in these cases, the correlation of problem one with problem two was calculated for those classrooms in which both problems were completed, the correlation was then corrected for restriction of range, and problem two scores were estimated by their regression on problem one, using the adjusted correlation. In the first sample, the same procedure was followed in predicting missing scores for problem three. Following this, the effect of grade level was removed statistically by calculating the correlation between grade level and total score, using this relationship to predict RSSR performance for each grade level, and taking the difference between predicted and obtained score for each classroom. These difference scores were then converted to standard scores for the total sample. In effect, grade norms were developed statistically, but based on the total number of classrooms in each sample. The same procedure was followed in providing the missing scores for problem two in the second sample.

Analysis

The first step in the analysis was that of processing the data by computer to obtain means, standard deviations, and intercorrelations. These were the basic data for most of the analyses which followed.

Criterion Composites

Since one of the core problems in defining teacher-effectiveness is the existence of multiple, essentially uncorrelated criterion dimensions, there is need for studying the means of combining these into composites. Three possibilities for meeting this need were explored in this study: 1. by canonical correlation, 2. by scaling by reciprocal averages, and 3. by averaging standard scores.

By Canonical Correlation. Canonical correlation is a procedure for creating a composite of predictor measures, and a composite of criterion measures, with weights assigned to each measure in each composite in such a way as to make the correlation between predictor and criterion composites a maximum. Three aspects of these results are of interest -- the nature of the composite of criterion measures (classroom behavior), the nature of the predictor composite (the personality and attitude measures of the teachers), and the extent to which the one is predicted by the other. This criterion composite is the one which is maximally predictable.

The computation was carried out by the available IBM 650 program, written on the basis of Hotelling's (1936) original article. The program is limited to four variables in one composite and ten in the other. The output of the program includes weights for each measure of each composite, and the squared correlation coefficient.

For the data of the first sample, canonical correlations were computed between the four criterion measures and the ten best predictors from all 24 measures, the ten best predictors from the MMPI alone, (since it appeared to contain most of the relevant variance), and the four best predictors from the MMPI.

By Scaling by Reciprocal Averages. Scaling by reciprocal averages is a procedure which weights each measure proportionally to its correlation with an initial a priori composite. Iterations are carried out until the weighting system stabilizes. The result is a composite which is maximally reliable, as contrasted to the maximally predictable composite created by canonical correlation. This procedure makes no reference to measures other than those which are being combined into the composite, in contrast to canonical correlation which creates two composites in relation to each other.

The description by Mosier (1946) was taken as an outline of procedure, but modified to be carried out by hand rather than machine. The modified procedure is outlined in Appendix C.

By Averaging Standard Scores. Averaging standard scores results in equal weighting of the variables being combined—the simplest procedure, against which the others can be compared.

Comparison of Composites. Similarities in the results produced by the three procedures were examined by comparing the intercorrelations of the composites with each other, with the best predictor measures, with the separate criterion dimensions, and by overlap of the extreme criterion groups selected by each for item analysis. Numbers of the usually expected tests of significance could not be applied for comparative purposes, because they do not exist. It would have been of interest to test whether ten MMPI keys, predicted the criterion composite better than four did, for example, but there is no test for significance of differences between canonical correlations (Hotelling, private communications). As another example, comparative reliabilities could not be computed for many of the composites,

because the differential weighting violates the assumption of homogeneity of variances required for an analysis of variance test. There is not, to the author's knowledge, a test for significance between multiple and canonical R. As a consequence, the problem of significance remains unexplored in many cases, and the only answers obtained are to questions of estimation.

Perhaps in an area where so little exploration has been done, failure to answer questions of significance does not negate the importance of the results. The questions of estimation still provide a beginning for examining the results of each procedure and their implications for teacher effectiveness.

Comparison of Canonical and Multiple Correlations. While canonical correlation would be expected to result in the highest interrelationship between several predictors and several criteria, the computation is extensive enough to make its application infrequent where computer facilities are not readily available. The alternative which is commonly employed and feasible without computer facilities, is multiple correlation, which maximizes the relationship between several predictors and a single criterion. In order to examine the question of whether canonical correlation results in a sufficiently large increase in predictive power to warrant the inconvenience involved in obtaining it, two sets of multiple correlations were run for comparison. One set was computed using the same predictors selected by the canonical correlation and creating the optimal weights for predicting each criterion measure in turn. The other set was computed taking the best predictors from those available for predicting each criterion in turn.